



## King's Research Portal

### Document Version

Publisher's PDF, also known as Version of record

[Link to publication record in King's Research Portal](#)

### Citation for published version (APA):

Woodfin, A. J., Kamenetzky, A., & Hinrichs-Krapels, S. (2017). *Collaborations among AMRC research-funding organisations: A preliminary analysis using Researchfish data: Summary of full report, 'The nature and scale of collaboration among AMRC research-funding organisations'*. The Policy Institute at King's. <https://www.kcl.ac.uk/sspp/policy-institute/publications/Collaborations-among-AMRC-research-funding-organisations.pdf>

### Citing this paper

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

### General rights

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Research Portal

### Take down policy

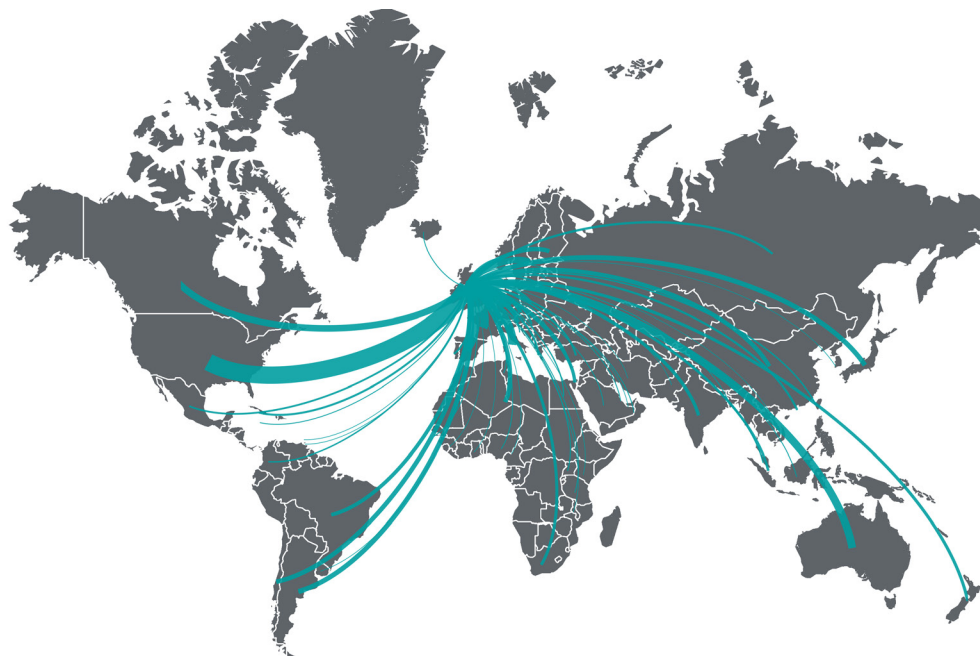
If you believe that this document breaches copyright please contact [librarypure@kcl.ac.uk](mailto:librarypure@kcl.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.

# Collaborations among AMRC research-funding organisations

A preliminary analysis using Researchfish data

---

Summary of full report, 'The nature and scale of collaboration among AMRC research-funding organisations', by Abigail Woodfin, Adam Kamenetzky and Saba Hinrichs-Krapels



Map of all global collaborations from AMRC organisations

**May 2017**

# Introduction

The Association of Medical Research Charities (AMRC) is the national membership organisation of leading medical and health research charities. Its members collectively fund over 40% of all publicly funded medical research in the UK. In addition to its role in setting high standards around the governance and administration of medical research funding, AMRC also provides guidance to its members, and advocates collectively on their behalf across the sector.

The Policy Institute at King's College London conducted a brief study to analyse the nature and scale of collaborations and networks across researchers funded by AMRC member organisations.

Specifically, the aims of this study were to:

1. demonstrate how data collected by AMRC member organisations to understand the outputs and impacts arising from their funding of research can be used to find collaborations, and
2. show the extent of any 'connectivity' between areas of research in the medical fields across charities.

To do this, we analysed the outputs of projects funded by AMRC member organisations using data collected within the Researchfish<sup>1</sup> platform. We also examined award and patent data from the Medical Research Council (MRC) and Wellcome Trust. This report is as much about demonstrating what is possible using Researchfish as an analytical tool as it is about showing those collaborations captured in this dataset.

---

<sup>1</sup> Researchfish is an online platform that enables research funders to capture and track the impact of their investments, and researchers to log the outcomes of their work. The platform has been designed to support researchers to report on the outputs (and outcomes and impact) of their work once, across multiple funders, and to reuse their data for their own use, and have control over who sees and accesses their data. The PI, or one of their delegates, can add, edit and delete output entries, and, crucially, attribute these entries to the research grants and awards that contributed to those outputs.

## Overview of data available

The number of principal investigators (PIs) that have projects funded by AMRC and are signed up to report on Researchfish total 3,355.

- These represent 255 research organisations (after cleaning).
- Between them, these PIs hold 5,287 awards within the Researchfish database, with a total value of £1,618,426,495.
- 9% of the PIs (n=289) recorded in this dataset hold awards from two AMRC funders, and only 1% hold awards from three or four funders; none hold awards from more than four funders.

We identified three ways in which collaborations and connectivity might be represented within this static snapshot of Researchfish data, using relatively simple 'high-level' analysis techniques:

1. **Reported partnerships:** These are instances of two institutions/organisations linked within Researchfish through PIs reporting them as partners within a collaboration.
2. **Co-authorship on publications:** These are instances of one or more PIs who have entered the same publication (identified by a unique PubMed ID)<sup>2</sup> as an entry under 'publications' and attributed it to one or more funding awards.
3. **Joint patents and licenses:** These are instances of one or more PIs who have entered the same patent or patent application (identified by a unique patent number)<sup>b</sup> as an entry under 'IP and licensing' and attributed it to one or more funding awards.

---

<sup>2</sup> Other types of outputs reported in Researchfish do not have unique identifiers and appear as free text submitted by PIs.

# Key observations

**The most encouraging finding was that we were able to identify the many geographical locations of partnerships and collaborations reported by AMRC member-funded researchers.**

This type of connectivity has never before been identifiable within any of the AMRC datasets. While roughly one half of collaborations were with other UK-based researchers, we identified over 500 institutions in 77 partner countries outside the UK with whom AMRC member-funded PIs reported a partnership (Figure 1).

Most of these collaborations are with the USA and Germany, followed by other countries in the European Union (which as a block represented 45% of all non-UK collaborations). Following cleaning and standardisation of PIs' institutional locations, a large number of locations were represented.

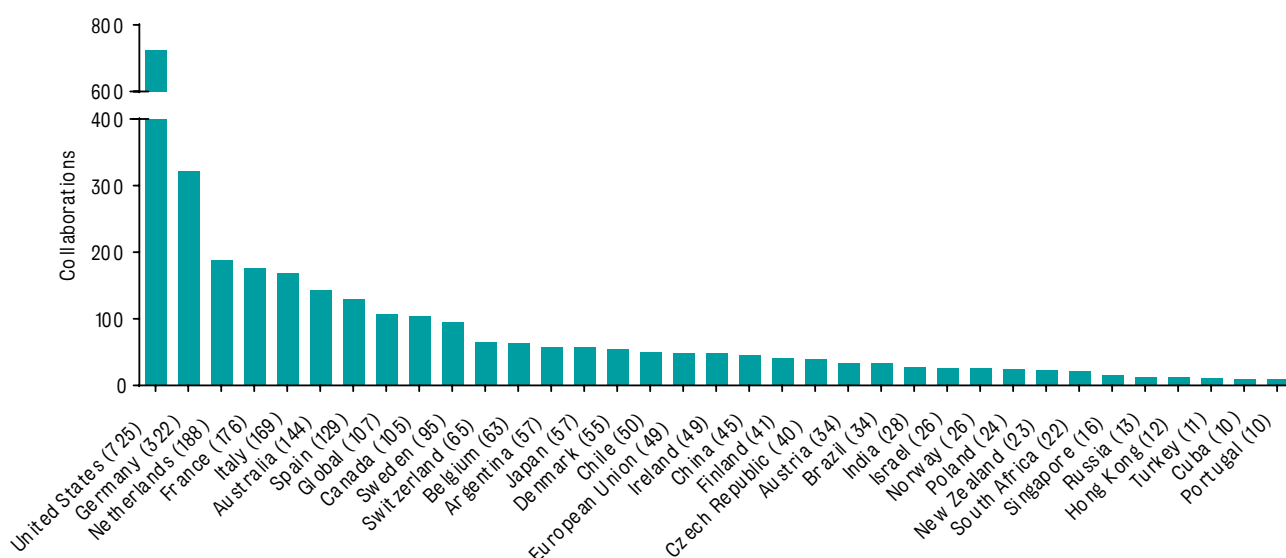
**Most (93%) of the 1,417 PIs who reported partnerships and collaborations are based in the academic/university sector, followed by hospitals (2%) and charities/non-profit organisations (2%).**

Just over half (54%) of these PIs' collaborating partners are from the academic/university sector, also followed by hospitals (13%) and charities/non-profit organisations (8%), as coded within the Researchfish dataset.

**The highest number of reported outputs captured within the Researchfish database – approximately one-half of the total – were publications, and we were able to identify linkages between funding organisations by using publications as a node of collaboration.**

Our analysis provides a visual representation of linkages between research-funding organisations and their grant holders, by using publications (specifically, those co-reported with common

**Figure 1:** Frequency of the countries listed as the location of the collaborating institutions; only countries with >10 reported partnerships are shown here

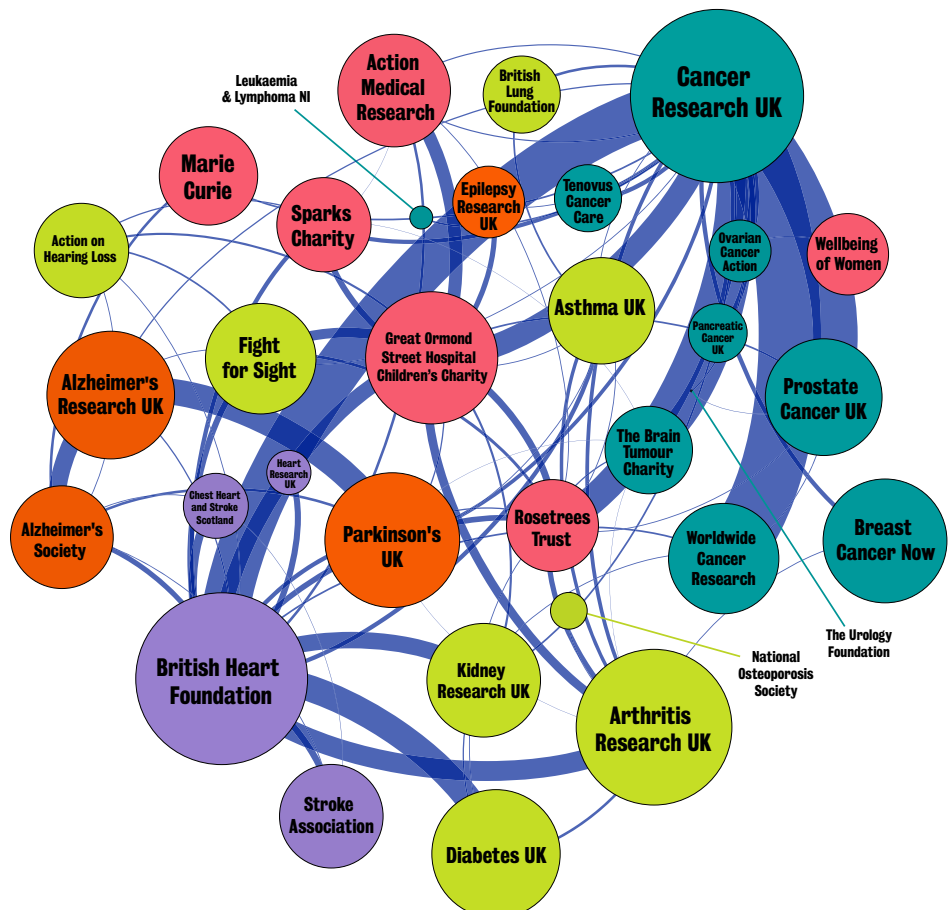


PubMed IDs to different funders) as a node of collaboration. This linkage map (Figure 2), while providing a vivid picture of the diverse network of connections across the AMRC funding ecosystem, nevertheless only represents a limited snapshot of true collaboration: only a subset of authors on any single publication are required to report outputs as PIs within Researchfish, and only a subset of these PIs will have attributed these outputs to different funders, to create a linkage in this fashion. We note that a majority of instances of co-reporting (82%) were by one or more PIs in receipt of funding from the same funding organisation, which may enhance individual funding organisations' visibility of their grant-holders' wider productivity. However, we estimate that only 1% of all publications reported through Researchfish represent what any single funding organisation might consider hitherto unknown instances of 'inter-funder' co-authorship (ie between different PIs holding awards from different funding organisations).

**Figure 2:** Linkage map of associations between different AMRC member organisations, derived from PIs co-attributing publications with the same PubMed IDs to different funding awards (i.e. accounting for duplicated outputs). [Funding organisation colour code: Pink – general medical research, teal – cancer research, lilac – cardiovascular research, orange – neurological research, lime – specific focus research] The node sizes indicate the value of awards held from each funding organisation (with an inverse log scale) and the thickness of the connections indicate the frequency of association of the funding organisations

The data for joint patents and patent applications are further examples of linkages between AMRC funding organisations.

We identified a total of 978 unique patent number or application IDs reported in the combined MRC, Wellcome Trust and AMRC Researchfish datasets. Co-reported patents attributed to awards from AMRC members (not including the Wellcome Trust) made up 4% of this total. The greatest number of these co-reported patents were between the MRC, and one of Parkinson's UK, the British Heart Foundation, and Arthritis Research UK. We found only two instances of single connections between AMRC member organisations themselves, namely between the British Heart Foundation and Diabetes UK, and between the cancer charities Ovarian Cancer Action and Cancer Research UK. As in each case the reporting was from one PI attributing an output to two funders, from this data alone, is not possible to determine whether the funders' dual support was part of an active collaboration between those funders, or the researchers combining multiple funding sources in support of their work. It will be of interest to see if this category of output increases as AMRC member-funded research moves through the process of translation.



# Caveats

## Data quality

We are only able to report on collaborations or partnerships in three ways, as described earlier (reported partnerships and collaborations, co-authorship on publications, and joint patents and licenses). For these three types of identified collaborations analysed, data received had to be matched and cleaned before it could be analysed. For example, 20% of AMRC and 9% of MRC journal articles did not have a PubMed ID. There was also significant variability in the format of patent application or ID number, so some joint work on patent applications may not have been identified. We therefore could not fully exploit the potential even in this dataset due to the required cleaning.

In addition to comparing with MRC data, we have also compared, where possible, with data received from the Wellcome Trust, but this was only possible for the publications analysis by identifying common PubMed IDs.

## Representativeness of sample

Reported collaborations and partners, publications and patents represent four of the 16 types of entries that can be captured through Researchfish. There is no standardised way to match the other output types. For example, two PIs may have contributed to the same policy documentation. These would appear as two separate entries and described differently despite referring to the same policy documentation (one may, for example, contain the exact document number to identify the documentation, while the other describes an abbreviated version of the title). Similarly, two PIs may have collaborated and contributed to a spin-out company, but only one of them names this company in its full form while the other simply describes the activities of the company in the free-text space. Furthermore, even some standardised reporting such as PubMed ID still does not capture items such as conference proceedings. Analyses of these types of outputs would be possible if such information was standardised, but the current state of the dataset does not currently allow for these.

Finally, our last caveat relates to what is not captured in the dataset. Every entry inputted into Researchfish by a PI is connected to a funding award (this is what enables linking across funders and analyses to take place across funders in a systematic way). However, we note that many informal collaborations also take place without direct funder support in an academic environment, and these would not be captured in our analysis.



# Reflections and learning on using Researchfish for this analysis

The Researchfish platform has been designed to support researchers to report on the outcomes of their work once across multiple funding organisations, reuse their data for their own use, and have control over who sees and accesses their data. In our previous report describing the potential for using Researchfish as an analytical tool,<sup>3</sup> we made a few observations for enabling funding organisations and research organisations to maximise the value of the data collected in Researchfish. These include the need to share data across funding organisations to enable informed comparisons across different outputs, to develop the analytical capability and capacity of funding organisations in using Researchfish data, to improve data integrity and quality by those who enter research outputs, and to ensure better connectivity of data within the research ecosystem with other data sources (such as publication databases and patent repositories). We frame our reflections in light of these recommendations:

**Sharing data across organisations to improve analytics.** Although we only had information for 40 AMRC member organisations available, we were able to generate unique visualisations of connectivity across funding organisations (eg using co-attributed publications as a locus of collaboration). The addition of a larger pool of data (eg the MRC portfolio) enabled us to see further instances of co-reporting of intellectual property exploitation, where this was based around unique IDs (such as patent/patent application numbers).

**Data integrity and quality.** The most frequently reported outputs in our dataset were publications. Although this may simply be a reflection of how academic researchers are used to reporting and/or incentivised, it may also be a reflection of the relatively recent start dates of the majority of funding awards, half of which clustered from 2012 to 2014. It was not within the scope of our

analysis to establish how accurately this dataset represents the full extent of PIs' outputs, or how PIs' self-reporting compares against automated data collection/harvesting from other sources (eg funder acknowledgements from bibliometric databases, and patent databases).

**Data connectivity.** In his review of the 2014 Research Excellence Framework, Lord Stern highlighted the need for interoperability between data systems used to collect metrics on research.<sup>4</sup> It is encouraging to see that Researchfish has begun to increase the connectivity of its data with other data sources in the research ecosystem. One such arrangement is with Europe PMC, which synchs the relevant funding organisation and award reference data both to and from Researchfish to display alongside publication data (effectively, providing an index of funding organisation-grant-publication matches). The MRC, and some of the larger medical research charities, purchase bibliographic data to generate a more complete list of co-authorships, which they then use to undertake further analysis (eg examining the proportions of authors from the private vs. charitable sector), providing a degree of triangulation. Further work is required to establish the representativeness of researcher self-reported publication data vs. automated data-harvesting methods (such as bibliometrics analysis of funder acknowledgements).

**Capability and capacity of funding organisations.** In terms of mapping broader connectivity across the UK charitable medical research-funding base, our analysis demonstrates that Researchfish data provides a novel view of partnerships and collaborations both within and outside of the UK. However, approximately one half of all international and UK partner institutions and organisations (45% and 55%, respectively) required a degree of formatting standardisation

<sup>3</sup> Hinrichs, S, Montague, E, Grant J. *Researchfish: a forward look* (2015) Policy Institute at King's College London.

<sup>4</sup> Stern, N. (2016) *Research Excellence Framework (REF) review: Building on success and learning from experience*. BIS, UK.

before geographical mapping could be carried out. There were also a large number of outputs with insufficient information to identify the partner organisation/institution. We are aware that since data was made available to us, Researchfish have put in place efforts to undertake this cleaning systematically, which should help AMRC member organisations to make better use of this data at the point of retrieval from the system.

Our analysis of patent data also revealed only limited capacity to identify novel instances of PIs collaborating across grants from different funding organisations (fewer than 4% of co-reported patents). Given the elapsed time that is to be expected for research to progress along a clinical pathway towards translation and commercialisation, it is encouraging to have identified even these small numbers, which indicate individual funders' support of research is being combined productively. However, as with our analysis of partnership data, cleaning was an issue: of 3,871 outputs with a patent license or application number from the AMRC, MRC and Wellcome Trust data combined, a total of 2,146 (55%) required reformatting or standardisation. We are aware of efforts to link Researchfish data with UK patent databases, and it would seem important to ensure that a route is provided to ensure this data is cleaned at, or immediately following, the point of entry, so that outputs can be usefully analysed and aggregated by AMRC and its member organisations.

Lastly, we note that there is an opportunity to improve the way Health Research Classification System (HRCS) data is utilised by AMRC member organisations. This would allow improved analysis and connectivity with respect to different health topics, as well as comparisons across funding organisations with similar remits. For instance, the Project Ecosystem study revealed that the British Heart Foundation (BHF) is acknowledged by a large number of authors of mental health research publications as a contributing funding organisation.<sup>10</sup> In this analysis, AMRC has chosen to use HRCS categorization of the award (input) as an indirect means to stratify outputs. Project Ecosystem showed that even for a funder that supports research in a focused area (eg cardiovascular), outputs could arise in other research areas. This may reflect a funder's desire to code according to their strategy, rather than the – potentially complex – realities of the research topic itself. Autocoding may provide a standardised route to HRCS coding, however this remains a relatively untested field.



[www.kcl.ac.uk/sspp/policy-institute](http://www.kcl.ac.uk/sspp/policy-institute)  
[@policyatkins](mailto:policyatkins@kcl.ac.uk)

---